JOURNAL OF BACTERIOLOGY, July 1996, p. 3860–3868 Vol. 178, No. 13
0021-9193/96/$04.00+0

Comparative Genomic Analysis of the *Haloferax volcanii* DS2 and *Halobacterium salinarium* GRB Contig Maps Reveals Extensive Rearrangement

ANDREW ST. JEAN AND ROBERT L. CHARLEBOIS*

Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

Received 2 February 1996/Accepted 19 April 1996

Anonymous probes from the genome of *Halobacterium salinarium* GRB and 12 gene probes were hybridized to the cosmid clones representing the chromosome and plasmids of *Halobacterium salinarium* GRB and *Haloferax volcanii* DS2. The order of and pairwise distances between 35 loci uniquely cross-hybridizing to both chromosomes were analyzed in a search for conservation. No conservation between the genomes could be detected at the 15-kbp resolution used in this study. We found distinct sets of low-copy-number repeated sequences in the chromosome and plasmids of *Halobacterium salinarium* GRB, indicating some degree of partitioning between these replicons. We propose alternative courses for the evolution of the haloarchaeal genome: (i) that the majority of genomic differences that exist between genera came about at the inception of this group or (ii) that the differences have accumulated over the lifetime of the lineage. The strengths and limitations of investigating these models through comparative genomic studies are discussed.

In recent years much effort has been devoted to genome-level analysis among prokaryotes (11, 15). Most of this effort involved the use of pulsed-field gel electrophoresis (PFGE), either in the construction of physical maps or in the direct comparison of restriction fragment patterns. The latter method allows for a large number of genomes to be quickly and relatively easily compared but suffers from poor resolution, although it has been useful in the typing of strains (18, 22, 32). A large number of studies have been conducted by using PFGE to construct physical maps (for a review, see reference 11). Most often, genetic markers are then localized to specific regions of the map through hybridization, allowing the investigation of gross rearrangements at the genomic level. Genetic loci occurring on the same restriction fragment, however, cannot be ordered on the map, masking any differences in their arrangement. This limitation restricts such comparisons to closely related genomes.

The majority of genomic comparisons performed to date have used PFGE-derived maps to compare genomes at the strain or species level. The results of these comparisons have prompted their division into two groups (15), organisms with highly conserved genetic maps and those with divergent maps. Examples of members of the former group include *Escherichia coli* and *Salmonella typhimurium* (40), *Borrelia* spp. (6, 36), *Clostridium perfringens* (3), *Lactococcus lactis* (28), *Mycoplasma* spp. (26, 37), and *Streptomyces lividans* (27). Members of the second group include *Bacillus* spp. (4, 5), *Rhodobacter* spp. (16), and *Leptospira interrogans* (48). The criteria for deciding in which group a comparison will be included have not been rigorously established, however. Typically, no objective measure is used to determine the degree of similarity between genomes, which makes relating the results of one comparison to those of another problematic. Indeed, genomes showing a moderate number of differences could be considered either conserved or divergent, depending on the context of the study.

* Corresponding author. Mailing address: Department of Biology, University of Ottawa, 30 Marie Curie, Ottawa, Ontario K1N 6N5, Canada. Phone: (613) 562-5800, ext. 6328. Fax: (613) 562-5486. Electronic mail address: robert@bio01.bio.uottawa.ca.

Of the more than 100 chromosomal maps available, only 10 are from members of the domain *Archaea*. To date, three archaeal genomic comparisons have been performed, one between two methanogens (46) and two between different members of the haloarchaea (19, 31). The most detailed maps among the *Archaea* are derived from the extreme halophiles. Maps of ordered cosmid libraries are available for *Haloferax volcanii* DS2 (8) and *Halobacterium salinarium* GRB (47), while highly detailed macrorestriction maps constructed by using two-dimensional gel electrophoresis are available for two additional strains of *Halobacterium salinarium* (19). The higher resolution of these maps compared to what is achieved with standard PFGE maps allows for more detailed and potentially informative comparisons to be made.

Comparison among the three *Halobacterium salinarium* strains showed almost complete conservation in the physical maps, with many identical restriction sites being found in all three. Only two variable regions were identified, the first spanning 240 kbp involving numerous changes to the restriction map, a large insertion-deletion, and an inversion and the second an insertion-deletion spanning approximately 10 kbp (19). While one would normally expect such conservation when looking at strains of the same species, the genetic instability of two of the strains in the comparison did not make this a forgone conclusion. Many strains of *Halobacterium salinarium* harbor active insertion sequences of different types in up to hundreds of copies (42). These elements are known to cause frequent insertional inactivation of chromosomal and plasmid genes (7, 12, 13). For some time, the genetic instability of *Halobacterium salinarium* was assumed to apply to the physical map as well (43). We now know that the map can be preserved despite the potential for rearrangement. Among members of the domain *Bacteria*, chromosomal rearrangements between repeated DNA sequences, both *rrn* operons and insertion sequences, are not uncommon (25, 29, 48).

A second comparison involving two halophilic *Archaea* of different species, *Haloferax volcanii* and *Haloferax mediterranei*, also found highly conserved maps (31). In this case, two inversions (one involving the two *rrn* operons found in this genus) and one transposition involving a single locus were found,

while the *Bam*HI restriction maps of the two chromosomes were completely different, as expected from interspecific sequence divergence. Because only a macrorestriction map exists for *Haloferax mediterranei*, some of the 35 probes used could not be ordered on its map. Within the resolution of the comparison, however, no other differences could be detected. This is despite the fact that *Haloferax volcanii* is known to possess active insertion sequences, although not in the same numbers as in some strains of *Halobacterium salinarium* (7, 44).

Because of the degree of conservation found in these two comparisons, we wished to discover if similar conservation applied to more distantly related halophilic *Archaea*. If the maps were conserved, an alignment would allow us to localize the homologs of cloned genes whose sequences might have diverged too much to be accessible through hybridization or PCR. Whether the maps were conserved or scrambled, the comparison would provide useful data for the study of forces which maintain or disrupt gene order (9). Of particular interest were the tempo and mode of genome-level change, set in a phylogenetic context. Such quantitation required implementing analyses which could provide a more objective measure of the degree of similarity between the genomes being compared than have been used in the past. The availability of detailed maps and cosmid libraries for the chromosome and plasmids of *Haloferax volcanii* DS2 and *Halobacterium salinarium* GRB made these two organisms the logical choice for this next haloarchaeal genomic comparison.

MATERIALS AND METHODS

Archaeal strains and cosmid libraries. DNA was obtained from the previously prepared ordered cosmid libraries of *Haloferax volcanii* DS2 (8) and *Halobacterium salinarium* GRB (47). These libraries included the chromosomes and the three largest plasmids found in each archaeon.

DNA dot blot and Southern blot preparation and hybridization. Dot blots were prepared by using the minimal cosmid libraries of *Haloferax volcanii* DS2 and *Halobacterium salinarium* GRB. Approximately 50 ng of each cosmid was mixed with an NaOH solution to a final concentration of 0.4 M and spotted onto GeneScreen nylon membranes (DuPont). For Southern blots used to verify ambiguous dot blot signals, cosmid DNA was digested with *Mlu*I, *Bam*HI, or a combination of both enzymes. Southern blots also used GeneScreen membranes and were prepared with approximately 0.5 μg of cosmid DNA per lane. DNA was transferred to the membranes by using a Tyler VT-20 vacuum transfer unit according to Tyler's protocol.

Hybridizations were performed as described in reference 47 with some minor changes. The prehybridization and hybridization temperatures were always 40°C, while the 1-h wash with 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–1% sodium dodecyl sulfate was done at 70°C. Most probes were fragments prepared from cosmid DNA digested with *Mlu*I, *Bam*HI, or both and isolated from agarose gels by using GeneClean (Bio101). Twelve genes cloned from various organisms (listed in references 8, 31, and 47) were also used as probes. In both cases probes were prepared by the random-priming method.

For higher-resolution mapping of the six pairs of probes that could not be ordered with the dot blots, Southern blots of the relevant cosmids plus flanking cosmids were prepared. Cosmids were digested with various combinations of one, two, or three of the following enzymes: *Bam*HI, *Bgl*II, *Dra*I, *Eco*RI, *Hin*dIII, *Mlu*I, and *Ssp*I for *Haloferax volcanii* cosmids and *Afl*II, *Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, *Mlu*I, and *Xho*I for *Halobacterium salinarium* cosmids. Each Southern blot was then hybridized with the appropriate pair of unordered probes. Partial restriction maps of the cosmids were prepared by using data from the digestions and the hybridizations so that the probes could be ordered on the genome.

Computer and statistical analysis. The computer program DERANGE II (provided by M. Blanchette and D. Sankoff, Université de Montréal) was used to determine the number of changes (inversions, transpositions, and inverted transpositions) necessary to transform one chromosome into the other given the order on each chromosome of a set of homologous loci. Data were run through DERANGE II with a variety of parameters: values of 4, 5 and 6 were used for "look ahead," while weights for transpositions and inverted transpositions were 1, 2, 2.5, 4, and 10. Values of 1 for inversion weight and 0 for length coefficients were used in all cases. The results were compared with values obtained by using 100 random permutations, and the significance level of the difference between the experimental and random results was calculated as 1 plus the number of randomized permutations having a "total cost" or "total number of moves" less than or equal to that for the experimental data, divided by 1 plus the number of randomizations. Total cost and total number of moves are values produced by DERANGE II that measure the degree of divergence between two DNA segments.

The performance of DERANGE II was tested with permutations containing known numbers of inversions. Successive random inversions were introduced into permutations of 35 loci until a total of 60 inversions had been done. Ten such sets of permutations were constructed. DERANGE II was then used to solve each permutation, and the results were averaged between sets. For this test, look ahead was set to 6 and transposition and inverted transposition weights were set to 2.5 or 10.

Conservation in the distances between loci on the two chromosomes was investigated by calculating every pairwise distance between the loci included in the test. These values were plotted, and a regression analysis was performed to look for any correlation between the two chromosomes.

RESULTS

Hybridization of probes to dot and Southern blots. The ordered cosmid libraries of *Haloferax volcanii* DS2 and *Halobacterium salinarium* GRB were used to prepare DNA dot blots representing the chromosome and the three largest plasmids of each genome: pHV4 (690 kbp), pHV3 (440 kbp), and pHV1 (86 kbp) for *Haloferax volcanii* and pGRB305, pGRB90, and pGRB37 for *Halobacterium salinarium*. *Halobacterium salinarium* cosmids were digested with *Mlu*I, *Bam*HI, or both enzymes, and selected anonymous fragments were used to probe the dot blots. Positive controls consisted of hybridization to the cosmid from which the probe was taken, and an equimolar amount of lambda DNA provided the negative control.

DNA for probes was taken from *Halobacterium salinarium* because it possesses fewer repeated sequences than does *Haloferax volcanii* (47), making interpretation of the results simpler. Also, there was concern that many probes from the 4.1-Mbp *Haloferax volcanii* genome would not hybridize to the 2.5-Mbp *Halobacterium salinarium* genome. A total of 143 anonymous probes ranging in size from 0.4 to 16 kbp were hybridized in this way, 120 from the chromosome of *Halobacterium salinarium* and 23 from its plasmids. Probes were as evenly distributed around the genome as possible (usually two probes per cosmid). In addition to these, 12 previously cloned genes were hybridized to both genomes. The dot blots allowed the hybridization signals to be localized to roughly a third of a cosmid, either the middle nonoverlapping portion, or within the regions of overlap with neighboring cosmids. This provided average resolutions of 15 kbp for *Halobacterium salinarium* and 14 kbp for *Haloferax volcanii*, depending on the sizes of the individual cosmids and the degrees of overlap with their neighbors.

Of the 143 anonymous probes that had been hybridized to the dot blots, 74 were found to give ambiguous hybridization signals (Fig. 1). To resolve these ambiguities, sets of Southern blots were prepared which included every cosmid showing an equivocal signal for a particular probe on the dot blots, plus positive and negative controls. We confirmed or refuted signals from 46 probes in this way.

In six instances, a pair of probes from different parts of the *Halobacterium salinarium* chromosome hybridized to a common locus. To dissect these signals, Southern blots of cosmids from the four unresolved loci on the *Haloferax volcanii* chromosome and the two on the *Halobacterium salinarium* chromosome were prepared. Single, double, and triple digests of the cosmids allowed partial restriction mapping of these cosmids, ordering three pairs of signals.

From the total of 155 hybridizations performed, 127 gave reliable results after the additional screening described above (Table 1). Of these, 70 probes cross-hybridized between the two genomes, including 12 gene probes. The sensitivity of the hybridization procedure used was tested by hybridizing a probe for ISH51 (the best-characterized insertion sequence family in

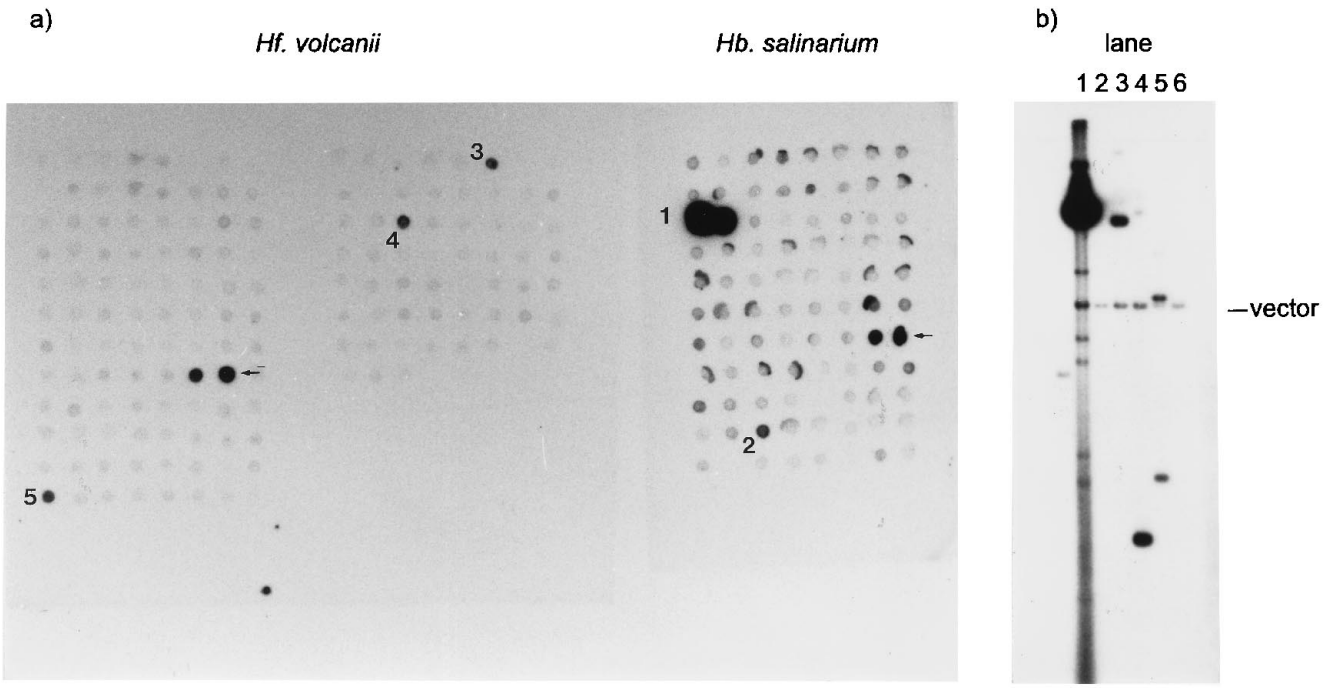

FIG. 1. Example of DNA dot blot hybridization of the *Halobacterium salinarium* and *Haloferax volcanii* genomic cosmid libraries giving ambiguous signals and the Southern blot used to resolve the ambiguities. (a) Dot blots were probed with a gel-isolated fragment of *Halobacterium salinarium* cosmid G19D7 giving a strong homologous hybridization in the overlap with an adjacent cosmid (numbered 1). Positive signals from two additional pairs of adjacent cosmids are shown with arrows; four weaker signals are numbered 2 to 5. (b) Southern blot of cosmids giving equivocal signals in the dot blots. The homologous hybridization is in lane 1, while the four questionable signals are in lanes 2 to 5. (Lane numbers correspond to the signal numbers in panel a.) Lane 6 is a negative control producing no signal on the dot blots. Three of the four ambiguous cosmids (lanes 3 to 5) produced signals on the Southern blot. Since the dot blot signal numbered 2 could not be reproduced on the Southern blot, we excluded it from the genomic comparison. The band to the left of lane 1 in panel b is a nonspecific hybridization signal to a lambda marker band.

Haloferax volcanii [10]) to the dot blots of both genomes (Fig. 2). No signals were observed for *Halobacterium salinarium* (strain GRB lacks the closely related ISH27 found in some other strains [39]), and all but one previously identified copy of the element (10) was found in *Haloferax volcanii*. Since ISH51 sequences can differ by at least 15% (20), this demonstrated that the procedure being employed could find homologous though divergent loci.

Repeated sequences. Twenty-nine probes gave multiple signals on one or both genomes (never more than five per genome) (Fig. 3). In a previous study using whole cosmids as hybridization probes (47), we identified five duplicated sequences in the *Halobacterium salinarium* GRB genome, four of them within or between plasmids. In that survey, hybridizations involving all cosmids representing plasmid DNA and cosmids representing 40% of the chromosome resulted in the conclusion that this strain's genome is quite repeat poor, in contrast to those of most other characterized strains. Here, our probes sampled 19% of the genome: 20% of the chromosome and 14% of the plasmid sequences. Ten of the 113 chromosomally derived probes hybridized to between two and five loci in the chromosome (none hybridized to plasmid DNA), whereas 6 of the 14 plasmid-derived probes hybridized to two or three plasmid loci and another plasmid-derived probe hybridized to a single chromosomal locus. We previously found that probes prepared from restriction fragments could be more sensitive than those prepared from whole cosmids (31); hence, the increased detection of repeated sequences is not surprising. A greater specific activity as well as a dissection of compound repeats may well be the explanation.

TABLE 1. Summary of hybridization results

Source of probe	Type of probe	No. of:		
		Probes	Unambiguous hybridizations	Unambiguously cross-hybridizing probes
Chromosome	Anonymous[a]	120	101	53
	Gene[b]	12	12	12
Plasmid	Anonymous[a]	23	14	5
Total		155	127	70

[a] Restriction fragments from cosmids representing the *Halobacterium salinarium* genome.
[b] Previously cloned from various haloarchaea (8, 31, 47).

Extrapolating from the present study, we can estimate the repeat content within the *Halobacterium salinarium* GRB genome to be approximately 50 low-copy-number repeats in the 2.03-Mbp chromosome and another 50 in the 0.43 Mbp of plasmids. Besides intragenomic repeats, 21 chromosomal probes and 1 plasmid probe hybridized to between two and four loci in the *Haloferax volcanii* DS2 genome (Fig. 3). Ten of these 22 probes gave multiple signals in both genomes. We do not know the nature of the repeated sequences; they may represent gene families as have been documented for the haloarchaea (1, 17, 21, 23, 41), insertion sequences (especially in the plasmids [7] though of uncharacterized types [14]), or other repeat sequence structures (34).

Cross-hybridizations. Seventy of the 127 probes giving clear hybridization results linked homologous loci between the two genomes (Fig. 3). It is likely that the 57 probes uniquely hy-

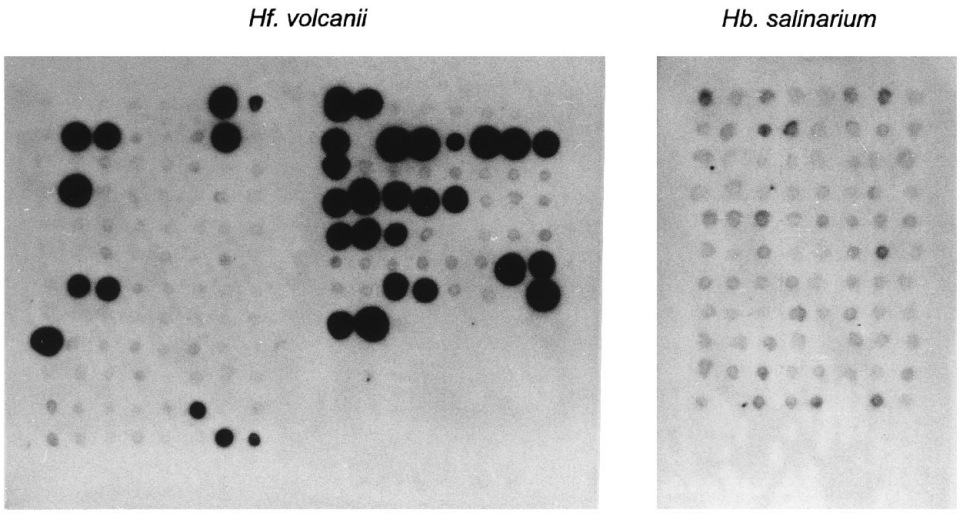

FIG. 2. Hybridization of the insertion sequence ISH51 to dot blots of cosmid libraries of the *Haloferax volcanii* and *Halobacterium salinarium* genomes. This insertion sequence does not occur in *Halobacterium salinarium* GRB.

bridizing to *Halobacterium salinarium* simply diverged from *Haloferax volcanii* in sequence beyond the threshold of detection, although major differences in genetic inventory between the two species cannot be ruled out. Of these 70 probes, 41 involved unique loci, whereas the other 29 included multiple signals (see above). Thirty-five of the 41 were interchromosomal, and 6 connected the *Halobacterium salinarium* chromosome to *Haloferax volcanii* plasmids (pHV4, three links; pHV3, two links; and pHV1, one link). Although the *Haloferax volcanii* plasmids make up 29.5% of its genome, only 17.1% of uniquely cross-hybridizing (18.5% of all cross-hybridizing) *Halobacterium salinarium* chromosomal probes found a *Haloferax volcanii* plasmid homolog. Most chromosome-to-plasmid, and all plasmid-to-plasmid, connections involved repeated sequences.

Analysis of comparison data. An overview of the chromosomal comparison led us to believe that extensive rearrangements had occurred but that some conservation remained in certain regions. The 35 probes producing one signal on each chromosome were used in analyses designed to quantify this impression.

The first analysis used the program DERANGE II (2, 24), which determines the minimum number of moves (using inversions, transpositions, and inverted transpositions) needed to transform one set of ordered loci into another. DERANGE II measures similarity in terms of total cost (where each move adds a set value to the total, with transpositions and inverted transpositions costing more than inversions) and number of moves (the total number of all moves needed for the transformation). Although DERANGE II can deal with circular DNA molecules, it inputs the order of loci as a linear set. In case this has an effect on the outcome of the analysis, each of the 35 circular permutations of loci was run through DERANGE II with the set of parameters listed in Materials and Methods. These results were compared with 100 randomized permutations of 35 loci (Fig. 4). Figure 4 shows results for total cost, which are similar to those given by number of moves (data not shown), indicating that there was no statistically significant ($P > 0.14$) difference between the randomized and experimental data.

In order to test how many changes between the two chromosomes were needed before DERANGE II would find no conservation, we had DERANGE II solve permutations of 35 loci with known numbers of inversions (Fig. 5). A linear relationship between total cost and the number of inversions is seen initially, eventually reaching a plateau. After roughly 40 steps, further inversions cease to have an impact on the ability of DERANGE II to solve the permutation. The total cost to solve random permutations of 35 loci closely follows this plateau, as does the total cost to solve the experimental data. Substituting number of moves for total cost gives identical results.

Another analysis of the experimental data involved determining whether conservation in the distances between pairs of loci within each chromosome exists. The position of a locus was estimated to be the center of the portion of the cosmid to which the probe hybridized. In a regression analysis (Fig. 6), no significant difference between the random and experimental data was found ($r^2 = 0.0148$).

DISCUSSION

There is no conservation in the order of loci on the chromosome or in their pairwise distances in the genomes of *Halobacterium salinarium* GRB and *Haloferax volcanii* DS2. Although a lower density of homologous connections between *Halobacterium salinarium* chromosomal DNA and *Haloferax volcanii* megaplasmid DNA than between the two chromosomes was observed, more than one in six chromosomal probes found a plasmid homolog. Genomic rearrangements which encompass all major replicons, shuffling loci between them, have occurred. The nature of plasmid-encoded loci is still unknown, however, except that few identified genes map to them (10). Probes hybridizing between plasmids and between chromosome and plasmids tended to confine themselves to FII and FI regions of plasmid DNA (8, 47), respectively.

Half of all cross-hybridizing probes found repeated sequences, numbering between two and five copies. These can include gene families, insertion sequences, and noncoding repeats. Insertion sequences concentrate in FII DNA (10, 38), and haloarchaeal plasmids are often FII (8, 47), but *Halobacterium salinarium* GRB is a genetically stable isolate (45) without any of the known *Halobacterium halobium*-type insertion sequences (14). Still, it is likely that the FII repeats are inser-

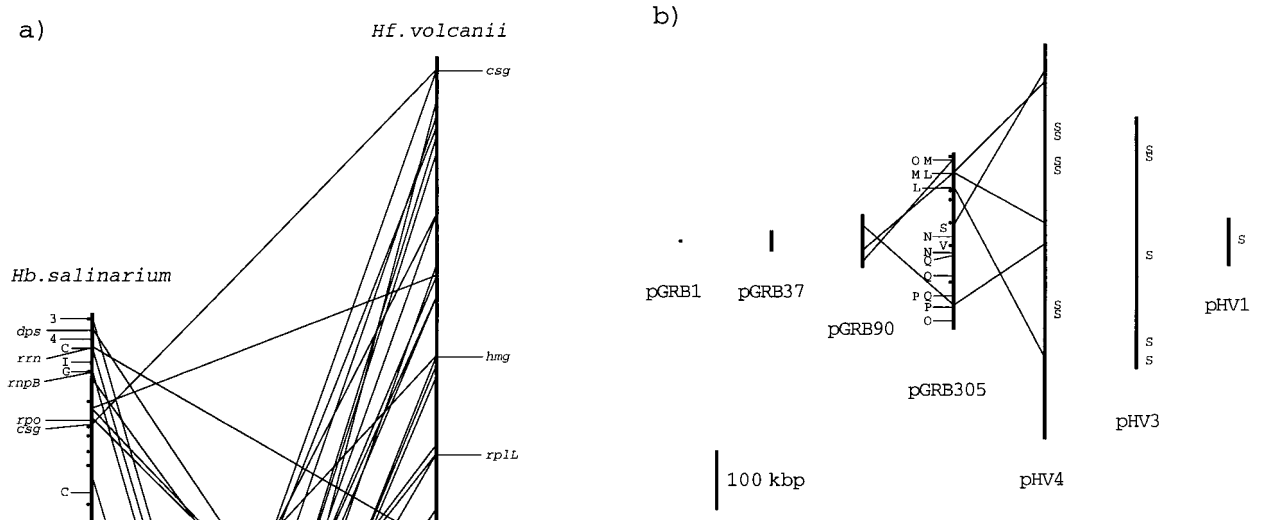

FIG. 3. Comparison of the *Halobacterium salinarium* and *Haloferax volcanii* genomes. All replicons are drawn to the same scale and are circular molecules represented as vertical lines for clarity. The top of each replicon corresponds to map position 0 (8, 47). Diagonal lines connect loci that cross-hybridize between replicons, while dots beside the vertical lines represent loci that do not cross-hybridize. All anonymous probes used were from *Halobacterium salinarium*. (a) The numbers 1, 3, and 4 are loci hybridizing to *Haloferax volcanii* plasmids pHV1, pHV3, and pHV4, respectively. Two loci hybridizing to the *Halobacterium salinarium* plasmid pGRB305 are indicated by 305. The letters A through K represent low-copy-number repeated sequences on the *Halobacterium salinarium* chromosome. Twelve gene probes are indicated by name. (b) The letters S and V represent loci hybridizing to the chromosomes of *Halobacterium salinarium* and *Haloferax volcanii*, respectively. The letters L through Q indicate low-copy-number repeated sequences on pGRB305.

tion sequences, albeit of new types, since haloarchaeal repeats are often insertion sequences (20) and insertion sequences cluster in FII (10, 38). We have observed the rare sectored colony of strain GRB, indicative of their presence and activity. We postulate that *Halobacterium salinarium* ancestrally maintains an adapted set of moderate insertion sequences but that the *Halobacterium halobium*-type strains recently inherited novel virulent types now wreaking havoc in their genomes.

Repeated sequences (especially insertion sequences) are often more conserved between haloarchaeal genomes than are unique sequences (39, 42). Our hybridization-based survey naturally demanded detectable, hence conserved, signals and may have overemphasized mobile genetic elements in the comparison and alignment. If strain GRB has been insulated from recent horizontal acquisition of novel insertion sequences, as evidenced by its lack of the most potent types, the multicopy signals are more likely members of older gene families.

Fifty-seven of 127 *Halobacterium salinarium* probes did not hybridize to the larger *Haloferax volcanii* genome. The lack of success of this 45% of the anonymous probes can result from major differences in genetic inventory between the species or from sequence drift beyond the threshold of detection by hybridization. This has consequences relevant to the interpretation of comparative genomic analyses such as the one presented here. Suppose, for instance, that a pair of duplicated genes map to parallel loci in two genomes. A probe including one of these genes would find both the paralogous and orthologous copies, revealing the duplicity. The hybridization may not give an indication of which copy is the ortholog, resulting in an inability to use that probe's results in a chromosomal alignment. If, however, sequence drift occurs such that only the intergenomic paralogous pair is detected by hy-

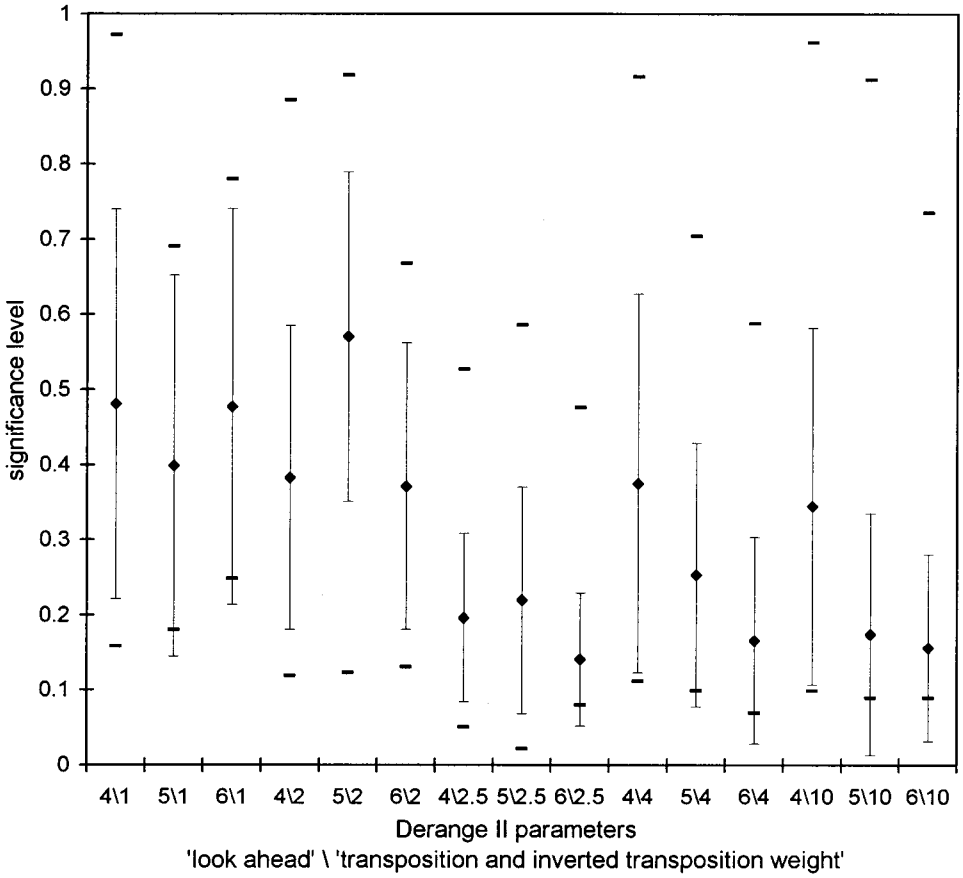

FIG. 4. Comparison between the total cost of rearranging 35 loci from the chromosomes of *Haloferax volcanii* and *Halobacterium salinarium* and random data. Numbers on the abscissa indicate the "look ahead" and "weight for transpositions and inverted transpositions" parameters of DERANGE II used in each run. Diamonds and vertical bars indicate average significance levels for the 35 circular permutations of the experimental data with associated standard deviations. Horizontal bars indicate the maximum and minimum values for each set of parameters used.

bridization (unlikely but possible if the copies are truly redundant and not under divergent selective pressures), a false genomic rearrangement is observed. Especially misleading and more common would be the cases in which alternate orthologs are deleted, leaving only a paralogous pair. Forty-five percent of our probes did not cross-hybridize, and 50% of those that did found multicopy loci.

Probes not cross-hybridizing were evenly distributed about the *Halobacterium salinarium* chromosome, and probes which did cross-hybridize found signals evenly distributed about the 44% larger *Haloferax volcanii* chromosome. Either there have been multiple small insertions or deletions in one genome relative to the other, or larger blocks have been scrambled by recombination. The latter scenario requires a more ancient origin of the chromosomal size difference. Although the *Halobacterium mediterranei* chromosome matches that of *Haloferax volcanii* in size (30), implying constancy since their divergence, a comparison within strains of *Halobacterium salinarium* (19) revealed recent small insertion-deletion events. Thus, neither mechanism can be excluded at this time.

There are now three haloarchaeal genomic comparisons available, at three phylogenetic depths: intraspecific, interspecific, and intergeneric. The first two demonstrated genomic stability, which is remarkable given the potential for disruption in haloarchaeal genomes. The present study found no amount of conservation in the order of loci. It is likely and expected that some assortments of genes have been maintained, especially operons. Macrorestriction maps were sufficient to align the genomes of *Haloferax volcanii* and *Haloferax mediterranei* (31), and they were ideal in the alignment of *Halobacterium salinarium* genomes (19). Here, we leap from an estimated interspecies divergence of 80 million years (31) to very roughly 600 million years, on the basis of a 16S rRNA divergence of 12% (33) and a clock rate of 1% per 50 million years (35). True rearrangement as well as complications involving gene families and sequence divergence (described above) effectively prevent chromosomal alignment, even with the use of high-resolution contig maps and a program like DERANGE II. It is important to remember that the *Halobacterium* and *Haloferax* genomes are derived from a common ancestor. Although extensive rearrangement, the full elucidation of which will require sequence-level comparison, was observed between them here, this rearrangement is a process driven by a balance of forces. What does it mean that these genomes have shuffled the order of their genes at a scale finer than our mean resolution of 15 kbp? Or indeed have they done so, given that we could have been misled by the dynamics of gene family sequence divergence?

We entertain two models of genomic restructuring: gradual and punctuated. The gradualist would see the process of rearrangement as a function of time. A more saltatory mode of genome evolution would see the genome abandon its map for another as a consequence of wholesale selection for an altered pattern of gene expression (9). Our present data, which are

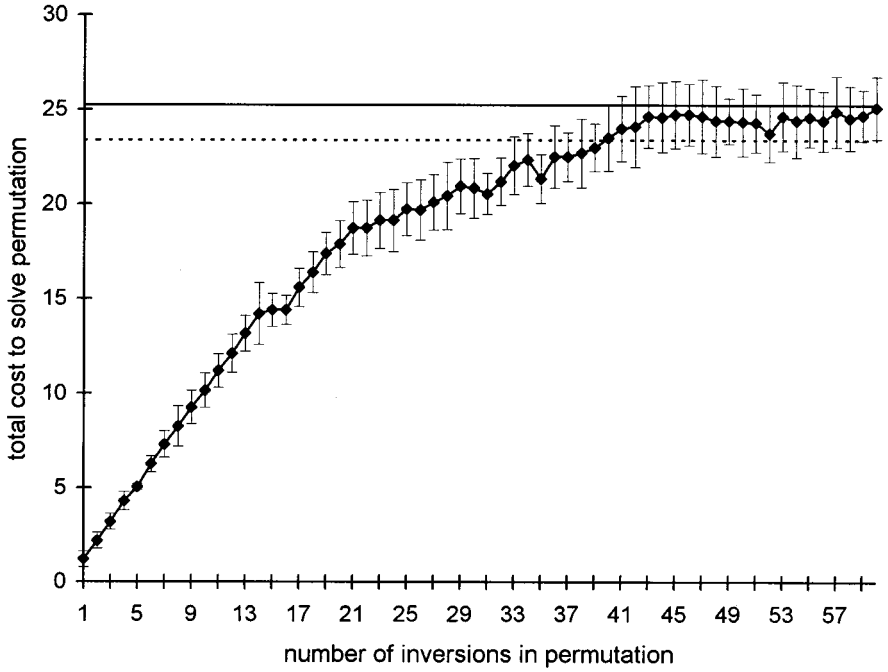

FIG. 5. Performance of DERANGE II on permutations of 35 loci containing known numbers of randomly generated inversions based on total cost. Permutations contain between 1 and 60 inversions. The DERANGE II parameters used were a "look ahead" of 6 and "transposition and inverted transposition weights" of 2.5. The solid horizontal line indicates the average total cost to solve 100 random permutations of 35 loci. The dashed horizontal line indicates the average total cost to solve the 35 circular permutations of the experimental data. Vertical bars indicate standard deviations.

limited to three comparisons, are consistent with both models. It is useful to address this issue of tempo, since rearrangement can effectively block recombination and thus contributes to speciation. Biodiversity, even among microbes, is a necessary component of survival and provides the foundation from which adaptations and innovations can arise.

If the haloarchaeal genome was designed at the base of their evolutionary radiation into (currently) eight genera through massive adaptive reorganization of an ancestral genome, there should essentially be eight maps. On the other hand, if rapid genomic reengineering did not coincide with the founding of the haloarchaeal lineage but rather has evolved to its present state through occasional rearrangements, there will be many maps which together may help us to construct detailed phylogenies and to recount the history of genomic events. The punctuated model would likely provide few clues in this direction, since the events would be condensed into a short span of time, and would effectively block a view beyond the origin to the earlier parent.

The lesson learned in this experimental study has been that, at least in the haloarchaea, genomic change is complex. Therefore, further studies should focus on fine detail in one or a few specific regions of the genome, with the assumption that similar things are happening elsewhere in the genome. Sequence-level comparisons can be used to measure the rate of genomic change relative to sequence divergence in many members of a lineage, although a reference sequence of an entire genome would be useful in order to solve the problem of paralogy.

Data need not be so extensive or so expensive as those obtained by genomic sequencing in order to answer many pertinent questions. The present study does, however, illustrate the need for tests to measure objectively the degree of similarity when genomic comparisons are performed. To date, most comparisons have involved low-resolution PFGE maps of

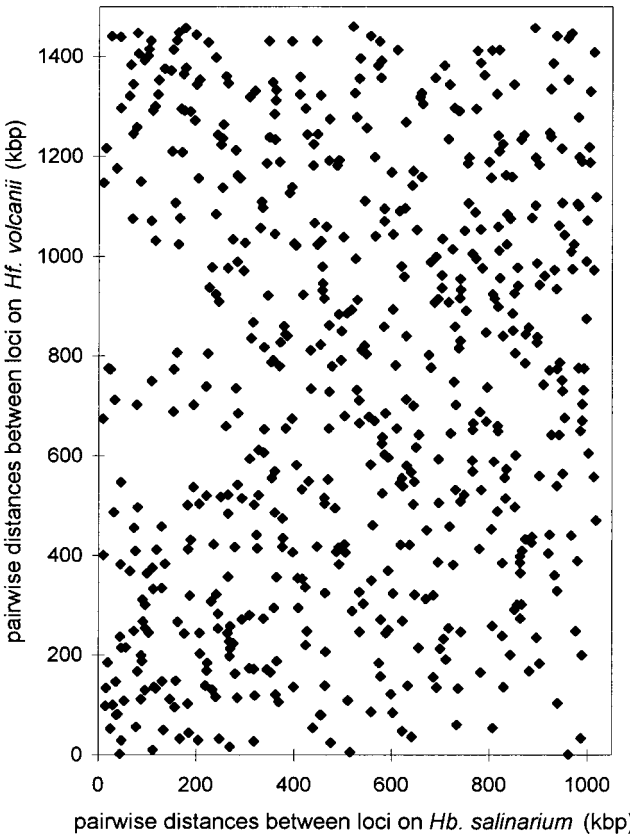

FIG. 6. Scatter plot of the distances between all pairs of 35 loci found on the *Halobacterium salinarium* chromosome and homologous pairs on the *Haloferax volcanii* chromosome. An r^2 value of 0.0148 was obtained by using 595 datum points.

closely related organisms, usually of the same genus or species. The small number of common loci used in these comparisons often makes the degree of similarity present easy to determine simply by looking at the data. Examples of comparisons in which many loci were used, such as those with *E. coli* and *S. typhimurium* (40) and *Rhodobacter capsulatus* (16), showed clear-cut results. Such easily interpreted results are unlikely to make up the majority of future comparisons.

The need for many cross-hybridizing signals in a comparison, at the highest possible resolution, is also clear. This becomes more important as the phylogenetic distance between the compared genomes increases, as shown by the present study, sometimes even necessitating large-scale sequencing. The existence of whole genome sequences of both prokaryotes and eukaryotes will encourage comparisons between distantly related genomes. These comparisons will have the most to gain from tests such as those performed in this study as well as future generations of increasingly sophisticated analytical tools. The move towards more objective measures and methodologies will be a necessary step if comparative genomic studies hope to answer any but the most vague questions about genomic level change and evolution.

ACKNOWLEDGMENTS

We gratefully thank D. Sankoff and M. Blanchette for providing the program DERANGE II. Assistance given by E. Weiher with the statistical analyses is also greatly appreciated.

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. R.L.C. is an Associate of the Canadian Institute for Advanced Research.

REFERENCES

1. **Antón, J., P. López-García, J. P. Abad, C. L. Smith, and R. Amils.** 1994. Alignment of genes and *Swa*I restriction sites to the *Bam*HI genomic map of *Haloferax mediterranei*. FEMS Microbiol. Lett. **117:**53–60.
2. **Blanchette, M., T. Kunisawa, and D. Sankoff.** Parametric genome rearrangement. Gene **172:**GC11–GC17.
3. **Canard, B., B. Saint-Joanis, and S. T. Cole.** 1992. Genomic diversity and organization of virulence genes in the pathogenic anaerobe *Clostridium perfringens*. Mol. Microbiol. **6:**1421–1429.
4. **Carlson, C. R., A. Grønstad, and A.-B. Kolstø.** 1992. Physical maps of the genomes of three *Bacillus cereus* strains. J. Bacteriol. **174:**3750–3756.
5. **Carlson, C. R., and A.-B. Kolstø.** 1993. A complete physical map of a *Bacillus thuringiensis* chromosome. J. Bacteriol. **175:**1053–1060.
6. **Casjens, S., M. Delange, H. L. Ley III, P. Rosa, and W. M. Huang.** 1995. Linear chromosomes of Lyme disease agent spirochetes: genetic diversity and conservation of gene order. J. Bacteriol. **177:**2769–2780.
7. **Charlebois, R. L., and W. F. Doolittle.** 1989. Transposable elements and genome structure in halobacteria, p. 297–307. *In* M. Howe and D. Berg (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
8. **Charlebois, R. L., L. C. Schalkwyk, J. D. Hofman, and W. F. Doolittle.** 1991. Detailed physical map and set of overlapping clones covering the genome of the archaebacterium *Haloferax volcanii* DS2. J. Mol. Biol. **222:**509–524.
9. **Charlebois, R. L., and A. St. Jean.** 1995. Supercoiling and map stability in the bacterial chromosome. J. Mol. Evol. **41:**15–23.
10. **Cohen, A., W. L. Lam, R. L. Charlebois, W. F. Doolittle, and L. C. Schalkwyk.** 1992. Localizing genes on the map of the genome of *Haloferax volcanii*, one of the Archaea. Proc. Natl. Acad. Sci. USA **89:**1602–1606.
11. **Cole, S. T., and I. Saint Girons.** 1994. Bacterial genomics. FEMS Microbiol. Rev. **14:**139–160.
12. **DasSarma, S.** 1989. Mechanisms of genetic variability in *Halobacterium halobium*: the purple membrane and gas vesicle mutations. Can. J. Microbiol. **35:**65–72.
13. **Derkacheva, N. I., V. K. Kagramanova, and A. S. Man'kin.** 1993. Genetic variability in halophilic archaebacteria (a review). Mol. Biol. **27:**287–295.
14. **Ebert, K., W. Goebel, A. Moritz, U. Rdest, and B. Surek.** 1986. Genome and gene structures in halobacteria. Syst. Appl. Microbiol. **7:**30–35.
15. **Fonstein, M., and R. Haselkorn.** 1995. Physical mapping of bacterial genomes. J. Bacteriol. **177:**3361–3369.
16. **Fonstein, M., T. Nikolskaya, and R. Haselkorn.** 1995. High-resolution alignment of a 1-megabase-long genome region of three strains of *Rhodobacter capsulatus*. J. Bacteriol. **177:**2368–2372.
17. **Gerl, L., and M. Sumper.** 1988. Halobacterial flagellins are encoded by a multigene family. J. Biol. Chem. **263:**13246–13251.
18. **Grothues, D., and B. Tümmler.** 1991. New approaches in genome analysis by pulsed-field gel electrophoresis: application to the analysis of *Pseudomonas* species. Mol. Microbiol. **5:**2763–2776.
19. **Hackett, N. R., Y. Bobovnikova, and N. Heyrovska.** 1994. Conservation of chromosomal arrangement among three strains of the genetically unstable archaeon *Halobacterium salinarium*. J. Bacteriol. **176:**7711–7718.
20. **Hofman, J. D., L. C. Schalkwyk, and W. F. Doolittle.** 1986. ISH51: a large, degenerate family of insertion sequence-like elements in the genome of the archaebacterium, *Halobacterium volcanii*. Nucleic Acids Res. **14:**6983–7000.
21. **Horne, M., C. Englert, and F. Pfeifer.** 1988. Two genes encoding gas vacuole proteins in *Halobacterium halobium*. Mol. Gen. Genet. **213:**459–464.
22. **Huber, I., and S. Selenska-Pobell.** 1994. Pulsed-field electrophoresis-fingerprinting, genome size estimation and rrn loci number of *Rhizobium galegae*. J. Appl. Bacteriol. **77:**528–533.
23. **Joshi, P., and P. P. Dennis.** 1993. Characterization of paralogous and orthologous members of the superoxide dismutase gene family from genera of the halophilic archaebacteria. J. Bacteriol. **175:**1561–1571.
24. **Kececioglu, J., and D. Sankoff.** 1995. Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement. Algorithmica **13:**180–210.
25. **Krawiec, S., and M. Riley.** 1990. Organization of the bacterial chromosome. Microbiol. Rev. **54:**502–539.
26. **Ladefoged, S. A., and G. Christiansen.** 1992. Physical and genetic mapping of the genomes of five *Mycoplasma hominis* strains by pulsed-field gel electrophoresis. J. Bacteriol. **174:**2199–2207.
27. **Leblond, P., M. Redenbach, and J. Cullum.** 1993. Physical map of the *Streptomyces lividans* 66 genome and comparison with that of the related strain *Streptomyces coelicolor* A3(2). J. Bacteriol. **175:**3422–3429.
28. **Le Bourgeois, P., M. Lautier, L. van den Berghe, M. J. Gasson, and P. Ritzenthaler.** 1995. Physical and genetic map of the *Lactococcus lactis* subsp. *cremoris* MG1363 chromosome: comparison with that of *Lactococcus lactis* subsp. *lactis* IL 1403 reveals a large genome inversion. J. Bacteriol. **177:**2840–2850.
29. **Liu, S.-L., and K. E. Sanderson.** 1995. Rearrangements in the genome of the bacterium *Salmonella typhi*. Proc. Natl. Acad. Sci. USA **92:**1018–1022.
30. **López-García, P., J. P. Abad, C. Smith, and R. Amils.** 1992. Genomic organization of the halophilic archaeon *Haloferax mediterranei*: physical map of the chromosome. Nucleic Acids Res. **20:**2459–2464.
31. **López-García, P., A. St. Jean, R. Amils, and R. L. Charlebois.** 1995. Genomic stability in the archaea *Haloferax volcanii* and *Haloferax mediterranei*. J. Bacteriol. **177:**1405–1408.
32. **Lück, P. C., J. H. Helbig, V. Drasar, N. Bornstein, R. J. Fallon, and M. Castellani-Pastoris.** 1995. Genomic heterogenicity amongst phenotypically similar *Legionella micdadei* strains. FEMS Microbiol. Lett. **126:**49–54.
33. **Mankin, A. S., V. K. Kagramanova, N. L. Teterina, P. M. Rubtsov, E. N. Belova, A. M. Kopylov, L. A. Baratova, and A. A. Bogdanov.** 1985. The nucleotide sequence of the gene coding for the 16S rRNA from the archaebacterium *Halobacterium halobium*. Gene **37:**181–189.
34. **Mojica, F. J. M., C. Ferrer, G. Juez, and F. Rodríguez-Valera.** 1995. Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. Mol. Microbiol. **17:**85–93.
35. **Ochman, H., and A. Wilson.** 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26:**74–86.
36. **Ojaimi, C., B. E. Davidson, I. Saint Girons, and I. G. Old.** 1994. Conservation of gene arrangement and an unusual organization of rRNA genes in the linear chromosomes of the Lyme disease spirochaetes *Borrelia burgdorferi*, *B. garinii* and *B. afzelii*. Microbiology **140:**2931–2940.
37. **Peterson, S. N., T. Lucier, K. Heitzman, E. A. Smith, K. F. Bott, P.-C. Hu, and C. A. Hutchison III.** 1995. Genetic map of the *Mycoplasma genitalium* chromosome. J. Bacteriol. **177:**3199–3204.
38. **Pfeifer, F., and M. Betlach.** 1985. Genome organization in *Halobacterium halobium*: a 70 kb island of more (AT) rich DNA in the chromosome. Mol. Gen. Genet. **198:**449–455.
39. **Pfeifer, F., and U. Blaseio.** 1990. Transposition burst of the ISH27 insertion element family in *Halobacterium halobium*. Nucleic Acids Res. **18:**6921–6925.
40. **Riley, M., and S. Krawiec.** 1987. Genome organization, p. 967–981. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
41. **Sanz, J. L., I. Marín, L. Ramirez, J. P. Abad, C. L. Smith, and R. Amils.** 1988. Variable rRNA gene copies in extreme halobacteria. Nucleic Acids Res. **16:**7827–7832.
42. **Sapienza, C., and W. F. Doolittle.** 1982. Unusual physical organization of the *Halobacterium* genome. Nature (London) **295:**384–389.
43. **Sapienza, C., M. R. Rose, and W. F. Doolittle.** 1982. High-frequency genomic rearrangements involving archaebacterial repeat sequence elements. Nature (London) **299:**182–185.

44. **Schalkwyk, L. C., R. L. Charlebois, and W. F. Doolittle.** 1993. Insertion sequences on plasmid pHV1 of *Haloferax volcanii*. Can. J. Microbiol. **39:** 201–206.
45. **Soppa, J., and D. Oesterhelt.** 1989. *Halobacterium* sp. GRB: a species to work with!? Can. J. Microbiol. **35:**205–209.
46. **Stettler, R., G. Erauso, and T. Leisinger.** 1995. Physical and genetic map of the *Methanobacterium wolfei* genome and its comparison with the updated genomic map of *Methanobacterium thermoautotrophicum* Marburg. Arch. Microbiol. **163:**205–210.
47. **St. Jean, A., B. A. Trieselmann, and R. L. Charlebois.** 1994. Physical map and set of overlapping cosmid clones representing the genome of the archaeon *Halobacterium* sp. GRB. Nucleic Acids Res. **22:**1476–1483.
48. **Zuerner, R. L., J. L. Herrmann, and I. Saint Girons.** 1993. Comparison of genetic maps for two *Leptospira interrogans* serovars provides evidence for two chromosomes and intraspecies heterogeneity. J. Bacteriol. **175:**5445–5451.